

Lorie A. Vanchena

## Reading German Culture, 1789–1918

- **Distant Readings/Descriptive Turns: Topologies of German Culture in the Long Nineteenth Century. 21<sup>st</sup> St. Louis Symposium on German Literature & Culture, Washington University in St. Louis, March 29–31, 2012.**

The 21<sup>st</sup> St. Louis Symposium on German Literature & Culture, »Distant Readings/Descriptive Turns: Topologies of German Culture in the Long Nineteenth Century«, organized by Matt Erlin and Lynne Tatlock, Department of Germanic Languages and Literatures, Washington University, took place March 29–31, 2012. Participants in the interdisciplinary symposium explored how the concept of ›distant reading‹ and its related technologies and methodologies could be used to study German literature and culture (1789–1918). Recognizing that digital technology provides scholars with access to a rapidly increasing amount of reading material and with new opportunities for searching and analyzing this material, speakers drew on the literary and cultural criticism of Franco Moretti, Stephen Best, Sharon Marcus, Robert Darnton, Wendy Chun, and others to consider »what can be gained (and what is lost) when we move away from an exhaustive rhetorical analysis of individual texts and turn our attention instead toward large bodies of data, making use of analytical techniques borrowed from such disciplines as statistics, computational science, quantitative history, and the emerging field of digital humanities«. <sup>1</sup> The speakers presented innovative research that explored a range of computational methods and tools, the relationship between distant and close reading, and the nature of reading itself.

In »Can Computers Read?« **Lutz Koepnick** (Washington University in St. Louis) identified the fears and desires that emerge when computing serves as a model or metaphor for reading. In Goethe's *Die Wahlverwandtschaften* (*Elective Affinities*), Eduard is upset by Charlotte's attempt to read a book over his shoulder; she obscures his personal engagement with the text and makes his reading a public act. Koepnick proposed that Eduard's response reflects today's »humanist fear« that computing may deny readers' autonomy by erasing differences between »semantic depth and textual surface«, while Charlotte's distant reading suggests the »post-humanist praise« of computers' ability to liberate reading from the »chimeras of romantic subjectivity«. How has digital computing redefined our understanding of reading and readers in society? What becomes knowable and what remains unknown when we view computational processes as objective and iterative, even if we do not understand them? Koepnick cited social theorists Luc Boltanski, Ève Chiapello, and Catherine Malabou, who define »the new spirit of capitalism« in terms of autonomy, self-control, and transparency – values that software and new media scholars Wendy Chun and David Golumbia, for example, claim shape our expectations for computers. <sup>2</sup> Although equating computing with human reading may result in new freedoms such as complete transparency, Koepnick warned that such freedom could come at a price: belief in the reader as »a self-reliant master over each and every text« can hide the degree of heteronomous control that exists in the invisible and incomprehensible. Can we theorize reading so that its »pleasure and promise« help readers defy the »neoliberal and entrepreneurial rhetoric of mastery« but still embrace the digital revolution and modern humanist culture? Koepnick proposed Freud's *Wunderblock* (›mystic writing pad‹) as a model for a kind of reading that interrupts readers' desire for control and opens their minds to wonder. Alberto Manguel argues that all writing needs a generous reader before it can acquire an active life, but Koepnick maintained that attributing computational reading with such generosity would be counterintuitive if we believe that computers have »metric, objectivist, and ra-

tionalist authority« that does not allow »the glitches, the detours and the subjective deviations« that bring texts to life for human readers.<sup>3</sup>

## Topology, Topic Modeling, and the German Novel

Speakers in Section I demonstrated that topological reading and probabilistic topic modeling can provide new insights into German literary history and the relationship between distant and close reading. **Andrew Piper** (McGill University), in »The Werther Effect: Topologies of German Literature, 1774–1832,« discussed his topological reading of Goethe's *Die Leiden des jungen Werthers* (*The Sorrows of Young Werther*, 1774). If *Werther* was a »syndrome«, he asked, to what extent did it influence other literary texts of that period? Piper described epistolary novels as »signs of a new culture of literary connectivity« characteristic of the commercial environment of eighteenth-century literature; as a fictional network of texts, the epistolary novel created actual textual networks. He and Mark Algee-Hewitt are building topological models to map the »lexical relationality« between *Werther* and Goethe's complete works. After modeling a »Werther« category (the ninety-one most common significant words), they measured the similarity of lexical redundancy in the corpus; the »Werther effect« is thus a system of repetitive differences. Piper's analysis revealed that *Werther* is fairly anomalous in Goethe's works; it correlates best not with its own period but with the revised version of the novel published in 1787. Furthermore, the notion of *Werther* as artist produces the strongest affinities between *Werther* and texts lexically similar to it. Piper also looked at the actional content of lexical relationalities in order to determine whether words in the *Werther* category can produce discourse: Wertherian words disappear in non-Wertherian clusters and words replace them, while other words persist. His approach revealed that the discourse on subjectivity and temporality typical of the epistolary novel yields to a new discourse on aesthetics. Piper then provided a valuable overview of the theoretical implications of reading topologically. Topology enacts a theory of social text, letting us study the interconnectedness of literary systems, for example. By shifting focus from the grammatical to the diagrammatical, topology also grants language dimensionality and allows for a vectoral reading of texts. Furthermore, topological models are spatial instruments that enable us to think temporally: the *Werther* effect structures our reception of the novel in the twenty-first century and makes us part of the effect. With its emphasis on linguistic redundancy and commonality, topological reading also redirects attention from lexical significance to the textual margins and thus helps us reconsider literary meaning.

**Matt Erlin** (Washington University in St. Louis) began »The Location of Literary History: Topic Modeling and the German Novel, 1731–1864« by suggesting that rethinking reading could lead to productive paradigms for our research other than the characteristic study of individual authors and discrete works. He posited that the concept of »distant reading«, conceived by Moretti and others as a turn away from conventional reading and toward scale and abstraction, as a shift from the individual text to the network, could help us develop new organizing frameworks. Drawing not only on the notion of distant reading but also on structuralist literary history and developments in humanities computing, Erlin sought to determine whether standard period designations for German literature (1750–1850) could be validated with data generated by probabilistic topic modeling. He also aimed to shed light on the »location of literary history«, i.e. on the commonalities used to create historically defined groups of texts. Topic models, developed in the field of machine learning and natural language processing to provide a statistical solution to the growing problem of managing electronic archives, are algorithms that analyze the words in large collections of documents in order to reveal not only major themes but also how the themes are connected and how they change over time.<sup>4</sup>

Applying topic modeling algorithms to his corpus of 154 canonical and lesser known novels (1731–1864), Erlin generated 100 topics (a topic being a list of terms likely to co-occur) and then determined the most common words for each topic and the extent to which each 1000-word chunk participates in a given topic. He found that the lists of topic words reveal a high degree of thematic coherence, which suggests that an automated classification of literary texts based on broad themes is indeed possible. Erlin also proposed that lists of topic words can help us understand authorial style as »a pattern of patterns«, for example, which moves us beyond similarity based on themes. Turning from topics to networks of texts, Erlin used network analysis to identify links between novels that topic modeling had clustered: the collection of novels constituted the network, and the similarities identified with topic modeling provided the links between pairs of novels. Erlin found that shared authorship seems to be the strongest indicator of similarity. He showed how topic modeling combined with network analysis can help us characterize an example of clustering based on shared authorship: his results revealed a relatively high proportion of adverbs present in Jean Paul's writings, which may reflect the author's »typically intrusive narrators«. In order to explain definitively this idiosyncratic aspect of Jean Paul's texts, however, we should return to the hermeneutic question of authorial style and to close reading; as Erlin observed, computers »cannot read *for us*«. Providing another illuminating example of how this »dialectical notion of the relationship between distant and close reading« can generate new avenues of inquiry, Erlin discussed the novels that constitute his »romanticism cluster«, the cohesion of which is driven predominantly by two love-related topics. A close reading of passages highly rated for these topics uncovered unexpected nuances relevant to current discourse on the gender of romanticism.

## Corpus- and Computer-Based Literary Analysis

Section II included discussions of available options for computational literary research. **Fotis Jannidis** (Universität Würzburg), addressing »Mapping the Narrative? A Corpus-Based Study of the German Novel from 1700 to 1900«, posited that corpus studies can complement traditional literary studies for which the text serves as the basic unit of literary history. Informed in part by Hayden White's work on the fictionality of historical discourse, his project seeks to examine »the ontology of (literary) history« by identifying the source of patterns in historical representation and by calibrating tools used for computer-based literary analysis. Using a corpus comprised of 350 canonical novels (1700–1900), Jannidis found that John Burrows's Delta, a method that uses frequent common words to measure the relative stylistic distance between texts as a way of testing authorship, to be a good indicator of authorial attribution for texts longer than 2000 words.<sup>5</sup> He also discussed R Script (Eder and Rybicki), which implements stylometric algorithms to measure elements of literary style; the script calculates the frequencies of words in the corpus and the most frequent words in individual texts and then determines a multidimensional distance for pairs of texts that can be depicted in bootstrap consensus trees, for example.<sup>6</sup> Jannidis's initial findings revealed potentially interesting anomalies: for instance, *Werther* does not cluster with Goethe's other works but with other romantic texts, and E. Marlitt clusters with male authors. One challenge with R Script, he pointed out, is relating the most frequent words to complicated concepts. If the project aims to reconstruct knowledge but at the same time challenge certain assumptions about genres, for example, when should work on calibration stop and the genre be redefined?

**Gerhard Lauer** (Universität Göttingen), speaking on »Calculating Literature: First Steps Toward a Computer-Based Analysis of Nineteenth-Century German Novels«, explored ways that mathematics can inform the study of literary history. A »new sociology of culture« is emerging, he observed, in part due to culturomics, an approach that creates massive datasets

and tools for the quantitative study »of human culture across societies and across centuries«.<sup>7</sup> Lauer cited recent research based on the millions of digitized volumes available in Google Books; Jianbo Gao, for example, studied the correlation between natural and social phenomena such as earthquakes and unemployment and concluded that word frequencies reflect how we approach different phenomena.<sup>8</sup> Lauer then reviewed computational tools available for studying nineteenth-century German novels. He showed how browsing Google Books for German novels published in 1809 identifies Goethe's *Elective Affinities* but also other novels not usually found in standard literary histories. Lexical analysis can be conducted using Voyant, a web-based tool for reading and analyzing digital texts that looks for word frequencies and enables one to consider how words such as ›love‹ or ›death‹ occur with other words. Principal component analysis, also using Voyant, creates word classes by putting words in the order of their frequency.<sup>9</sup> Using Stylometry with R (Eder and Rybicki), Lauer uncovered a »kinship« between novels such as *Elective Affinities* and Stifter's *Nachsommer* (*Indian Summer*) – a result reached, he emphasized, »not by reading«. He proposed that eAQUA, which applies text mining technologies to ancient texts, could also be used to study Bible quotations in *Effi Briest*, for example.<sup>10</sup> Whether closely analyzing one text or determining how reuse of the Bible changes over time, results obtained with eAQUA provide a cultural view of literary history. Lauer suggested important future steps for computational literary analysis, including more comparative studies, topic maps and sentiment analysis, and research on narrative features of texts.

## Distant and Close Reading

Presenters in Section III considered both positive and negative implications of distant reading and its applications to German literary and cultural history. The project **Tobias Boes** (University of Notre Dame) introduced in »The Vocations of the Novel: Distant Reading Occupational Change in Nineteenth-Century German Literature« utilizes a database containing about 11,000 book-length works of German-language prose fiction (1750–1950). Keywords he assigned to each text designate professions that receive extended narrative treatment, so users can track depictions of professional life over time and formulate hypotheses about the relationship between the novels and social change. Among the fifteen large vocational clusters Boes identified in novels published 1848–1919 are the agricultural and artisanal professions, the arts, clergy, media, and health. His initial data showed little obvious correlation between the vocations in the novels and real life. Clergy, for example, did not constitute 15% in terms of social significance, as they did in the corpus, but as Boes suggested, priests made good fictional characters in certain types of novels. Literary depictions of clergy sharply declined from 1871 through the 1880s, during the *Kulturkampf*, but the number of Protestant figures actually increased while the number of Catholics decreased. Scientific labor does correlate statistically with real life, as evidenced by a sharp increase in such professions after 1870. Shifting to a more speculative literary history, Boes proposed that his data could help us ascertain whether some forms of the novel are predisposed to certain vocational depictions. He finds it problematic to submit traditional genres to abstract analysis, as Moretti does in *Graphs, Maps, Trees* when discussing the life cycles of literary genres.<sup>11</sup> Furthermore, while some types of surface reading (determining the number of books published by year, for example) constitute easily quantifiable forms of distant reading, Boes called for new models that integrate numerical descriptions and expand the spectrum to include close and large-scale distant reading and all points in between.

In »*Black Devil and Iron Angel* Revisited: N-Gramming the Railway in Nineteenth-Century German Fiction«, **Paul Youngman** (UNC Charlotte) reminded us that information technol-

ogy has been challenging and changing the humanities since the invention of the computer in the 1940s. He proposed a »moonshot« digital humanities project similar to the Google Books ngram viewer: curating and expanding the German corpus currently available in Google Books (37 billion words), an undertaking that would necessitate not only a shift toward quantitative, explanatory research but also considerable resources and »massive collaboration«. As Youngman noted, computers are capable of performing tasks on a large cross-section of books that can enhance (but not necessarily replace) traditional methods by allowing us to identify patterns and conduct analyses based on plot or syntax, for instance – just the types of questions we have always tried to answer. Unlike Moretti, Youngman argues that interpretation and explication are similar: both involve identifying and interpreting patterns. Indeed, the ngram viewer confirmed several claims he made in *Black Devil and Iron Angel: The Railway in Nineteenth Century German Realism* (2005).<sup>12</sup> Results of a search for »Eisenbahn« in books published 1835–1900, for example, support the trajectory of the railway as a cultural trend that peaked around 1871. The tool also confirms the centrality of the railway through 1900 relative to the telegraph, loom, and steamship. The ngram viewer shows that Berthold Auerbach was a fairly consistent trend in the nineteenth century; somewhat surprisingly, he trended higher than the canonical authors Youngman had studied (Hauptmann and Fontane) from 1870 until about 1912. Given how the Realist authors trend well into the twentieth century, with most of them peaking after they had died, Youngman suggested that it would be interesting to consider whether canon formation influenced such trends. He emphasized that the ngram viewer does not offer conclusions. Applying quantitative methods to literature is not new, but Google Books and culturomics are, as is the scale they offer humanities research »in the name of getting things less wrong«.

»The Case for Close Reading after the Descriptive Turn« made by **Todd Kontje** (UC San Diego) reflected his skeptical but not dismissive view of distant reading. He noted the impact of Moretti's concept of distant reading on literary studies, including renewed interest in the book and its institutional and social contexts as well as work by Stephen Best and Sharon Marcus on surface reading and by Heather Love on the descriptive turn and »close but not deep« reading.<sup>13</sup> Although the digital revolution has dramatically changed the way literary scholars access and disseminate texts, Kontje urged us not to abandon »the slow pace of close reading«. He finds literary genres more complex and their evolution »considerably less mysterious« than Moretti claims: the *Bildungsroman*, for example, has been called the typical form of the nineteenth-century German novel but also the »missing or phantom genre«. Close reading, not »the mysterious rhythms of formal innovation«, Kontje suggested, can perhaps best explain why some novels resist generic labeling. Form is slowly taking precedence over content in Moretti's work, Kontje observed, with the »historical and cultural specificity« of novels being replaced with abstraction and »seemingly scientific objectivity«. Expressing concern for literary works on the margins (novels by nineteenth-century German women writers, for instance) that were excluded from the canon for reasons more ideological than qualitative, he advised against a »dogmatic insistence on distant reading« as the only means of revising the canon; here he disagrees with Moretti, who proposes that all literary works be studied and does not acknowledge good reasons for excluding some from the canon.<sup>14</sup> Kontje also connected the ongoing debate about publishing online as the first step toward scholarly exchange (as opposed to publishing the final product in print) to the discussion of distant reading: the close reading needed to create critical editions is »diametrically opposed to distant readings of large databases«. As he pointed out, however, close reading of one text does not preclude distant readings of other texts, just as distant readings do not preclude critical editions. Finally, he cautioned that digital maps and graphs might create a »misleading sense of pseudo-scientific objectivity in the humanities«. While distant readings can create new modes of inquiry, Kontje encouraged us to shift our focus periodically »to objects closer at hand«.

## Detoured Reading

**Jonathan Hess** (UNC Chapel Hill), who was to speak on »Distant Reading and the Study of Nineteenth-Century German-Jewish Culture« in Section IV, was unable to attend the conference. **Katja Mellmann** (Universität Göttingen), in »»Detoured Reading<: Understanding Literature through its Contemporary Reception. Case Studies in Nineteenth-Century German Novels«, proposed »detoured reading«, the effort to locate and analyze commentary that reveals how a work was perceived by its initial readers, as a type of distant reading. Literary historians seldom take this approach, she maintained, as evidenced by the relative lack of critical editions of original reception. The historical record may be incomplete, but reception analysis can provide an »empirically true« reading and help prevent anachronistic analyses. Mellmann has found that early reviews (1868) of E. Marlitt's *Goldelse* (*Gold Elsie*) and *Das Geheimniß der alten Mamsell* (*Old Mam'selle's Secret*), which were serialized in *Die Gartenlaube*, are critical of the novels' anti-religious tendencies. According to Mellmann, the perception of Marlitt's works as well-written romance novels did not explain their initial success, which she attributes instead to their engagement with the liberal tendencies typical of Keil's periodical – an aspect of Marlitt's writing neglected by previous scholars. Mellmann has undertaken a detoured reading of Freytag's *Soll und Haben* (*Debit and Credit*) that focuses on whether the novel was perceived as anti-Semitic by its initial readership. She has found that some reviews of early editions criticize Freytag's negative portrayal of Jews and that all acknowledge but do not necessarily condone how the author contrasts Jews, noblemen, and Poles with the more favorably depicted German middle class. Drawing on Niklas Luhmann's theory of socio-cultural evolution to address critics' claim that *Soll und Haben* resonated among readers with anti-Semitic inclinations, Mellmann has determined that Freytag's novel, at least in the first decade after its publication, did not directly generate changes in the contemporary cultural discourse.

## Distant Reading and Transnational Culture

In Section V, speakers used the concept of distant reading to provide new insights into transatlantic cultural transfer and literary history. **Kirsten Belgium** (University of Texas at Austin), in »Distant Reception: Bringing German Books to America«, explored the »particularly foreign encounter« that occurred in the early nineteenth century when American intellectuals visited Germany and its libraries and returned home with books. Inspired by Moretti's description of distant reading as a way of obtaining »a sharper sense of [the] overall interconnection« among texts, Belgium proposed »distant reception« as a new critical framework to complement traditional close reading.<sup>15</sup> Drawing on Robert Darnton's observation, »Statistics do not tell a story by themselves, of course, but they can open the way to various narratives by revealing patterns«, Belgium analyzed bibliographic lists in order to study the role of foreign books in American culture.<sup>16</sup> She demonstrated that a broader perspective on German influence in American libraries, one that encompasses publishing and the ways in which books were disseminated and collected, illuminates the complex nature of international cultural exchange, in this case German ideas and scholarship in America. Books in German were seldom found in American collections or libraries before 1830. In 1823, however, the Yale College library contained a significant number of works published in Germany (most in Latin) in the fields of theology, classical languages, and natural history. Joseph Stevens Buckminster, Jr., pastor of the Brattle Street Church in Boston, did not own any German fiction, but his library included some translations of German non-fiction as well as 107 books published in German cities (again, most in Latin) on classical and biblical antiquity. The patterns Belgium discovered in the records of these and other collections underscore American interest in German

intellectual life in the early 1800s. As Belgum observed, the evidence of Americans who studied antiquity, theology, and natural history using books published in Germany also shows that cultural transfer does not necessarily occur between national cultures.

Building on Darnton's concept of the ›communications circuit‹ and recent approaches to book history, **Lynne Tatlock** (Washington University in St. Louis), in ›The One and the Many: *The Old Mam'selle's Secret* and the American Traffic in German Fiction (1868–1917)‹, considered E. Marlitt's novel in relation to other American books and their publishers and readers, illustrating how popular literature circulated ›across political and linguistic boundaries over time under certain industrial conditions‹.<sup>17</sup> A close reading of *Secret*, first published in America in 1868, reveals that the domestic romance reflects German middle-class values and views of the ›cultural nation‹ as well as elements typical of the German domestic fiction then popular in America. International influences on Marlitt's writing also surface: *Secret* bears a resemblance to Brontë's *Jane Eyre*. Turning to different modes of distant reading, Tatlock presented publication statistics on 103 American issues of *Secret* (in three American translations) and 67 exemplars (1868–1926); her extensive data confirm the novel's position as ›the longest-enduring example of nineteenth-century German domestic fiction in American translation‹. Furthermore, records from the Muncie Public Library and signatures and/or dedications in the exemplars evidence historical readers: *Secret* had a largely female readership, and Marlitt was the tenth most widely circulating author in the Muncie library 1891–1902.<sup>18</sup> Tatlock also found that German heritage did not play a significant role in readers' preference for the novel. Using a publishers' survey and topic modeling, Tatlock showed that the books American publishers considered marketable in 1876 had word affiliations similar to those in *Jane Eyre*, although *Secret* correlates more closely with German domestic fiction.<sup>19</sup> Applying topic modeling to the three translations of *Secret* produced similar results, despite differences in the language used by each of the translators. As Tatlock suggested, topic modeling, although based on linguistic collocations, can identify affinities ›at a deeper level than word choice‹. Tatlock's broad approach to transnational literary history also allows a rethinking of Moretti's claim, in *Graphs, Maps, Trees*, that the life span of ›normal literature‹ is twenty years. Tatlock pointed out that Moretti counts new titles only, without considering how translation, reprinting, and new audiences extend the lifetime of a literary work. *Secret*, through republication and new marketing strategies, remained popular for about forty years.

## Case Studies, Periodicals, and Journal Articles

Speakers in Section VI offered illuminating examples of distant reading applied to archival fiction, nineteenth-century periodicals, and literary criticism. **Nicolas Pethes** (Ruhr-Universität Bochum) maintained in ›Serial Individuality: Case Study Collections around 1800‹ that early nineteenth-century authors recognized the influence being exerted on their era by a new mass market for weekly and monthly periodicals and the accompanying demand for shorter texts. Pethes suggested that these historical circumstances call to mind Moretti, who in *Graphs, Maps, Trees* replaces individual works with large sets of data. Pethes focused on the case study, a genre he finds well suited to quantitative research. He argued that a distant reading of case studies leads us to ›archival fiction‹, in which metaphors for archives, for instance, reflect these new market conditions. Rather than apply quantitative methods to literary analysis, Pethes looked at how literature suggests a quantitative perspective; quantitative research becomes a tool for and the result of distant reading. Empiricism in the new human sciences, Pethes observed, had resulted in archives of case studies, and writers during the Age of Enlightenment and the nineteenth century embraced the idea of ›serial individuality‹, the notion that an individual's narrative existed within a larger archive. He provided two exam-

ples. In Goethe's *Wilhelm Meisters Lehrjahre* (*Wilhelm Meister's Apprenticeship*), the scroll that Wilhelm receives from the Tower Society bears similarities to a case study, and the main characters all have case files in the Tower's chapel. The briefcase in Stifter's *Die Mappe meines Urgroßvaters* (*My Great-Grandfather's Briefcase*) is a book of medical case studies that also constitutes the work's narrative structure; Stifter thus presents the novel, published in the periodical *Wiener Zeitschrift*, as »from, in and as a case archive«. By connecting empirical medicine, case histories, and the need to organize and analyze the data they produce, *Mappe* finds echoes in Moretti's call for quantitative analysis of literature. As Pethes observed, however, Stifter, unlike Moretti, calls attention to the historical construction and contingency of quantitative research conducted on large archives of text.

In »Rethinking Non-Fiction: A Digital Humanities Approach to the Nineteenth-Century Science-Literature Divide«, **Peter McIsaac** (University of Michigan) illustrated how strategies of distant reading can help scholars approach the vast amount of material in nineteenth-century German periodicals. Such strategies, he proposed, can also generate new questions about changes in editorial configuration over time and shed light on the interaction between »particularities of nineteenth-century culture« and the changing publishing landscape. Studying *Die Gartenlaube* and *Deutsche Rundschau* in a holistic manner, McIsaac mapped digitized article indices synchronically and diachronically. His data show that both journals experienced shifts in their respective configurations that challenge accepted accounts of the periodicals. For example, whereas Keil claimed that the popularity of the *Gartenlaube* rested on its national project and its mission to popularize scholarly and scientific knowledge, the natural sciences and medicine constituted a far smaller share of the journal's content after the initial years. McIsaac suggested that this might be explained by lingering perceptions that readers had formed during the journal's early period, which were then reinforced by the republication of nonfiction texts as books and brochures in the 1850s and 1860s. In the *Rundschau*, the amount of serialized fiction varied much more than previously thought, and it was generally as low as 26%, despite the leading role attributed to literature in the journal's stated program. In both periodicals, moreover, McIsaac found that the relationship between fiction and non-fiction changed around 1885. In the *Gartenlaube*, serialized literature surged relative to historical/descriptive articles; in the *Rundschau*, non-humanities disciplines and literature tended to become decoupled as concentrations emerged for literature-humanities and science-history-politics. As McIsaac demonstrated, the journals' identities and their readers' interests need to be modeled in more dynamic, holistic ways; even simple data can reveal complex interactions between the publishers' stated programs, editorial configurations, and republication practices.

**Allen B. Riddell** (Duke University), in »How to Read 16,700 Journal Articles: Studying Nineteenth-Century German Studies Using Topic Models«, submitted that machine reading in general, and topic modeling in particular, offer practical ways to study more journal articles than we can read. Riddell based his analysis on a corpus of over 20,600 articles from *Monatshefte*, *The German Quarterly*, *New German Critique*, and *German Studies Review*, which he created using JSTOR's publicly-available Data for Research service. Riddell applied topic modeling algorithms to his corpus in order to identify trends in the history of German Studies. Discussing some of the theory behind topic modeling, he explained how topic modeling builds on earlier approaches to text clustering that used vector analysis. In vector analysis, texts are reduced to lists of terms, or »bags-of-words«; using word frequencies, these lists allow us to represent a text as a vector in a multidimensional space. If, for example, in chapter 1 of *Effi Briest*, Effi is mentioned 21 times and Innstetten 7, then 21,7 becomes a vector that can be represented using the vector space model. By reducing a corpus to a group of such vectors, we can measure the similarity and dissimilarity between texts by calculating the co-



sine distance between any pair of vectors. Riddell made clear that vector analysis has its shortcomings, some of which topic modeling attempts to address. Topic models identify not only individual word frequencies, but also clusters of terms that tend to co-occur in a given collection of documents. Topic models also have the advantage of letting us distinguish between different uses of the same word, such as the ›bank‹ of a river and the ›bank‹ in which one saves money. This level of sophistication is achieved by applying a complex statistical model, Latent Dirichlet Allocation (LDA), which introduces probability into the measurement of similarities among texts. While LDA also has shortcomings, it offers excellent possibilities for ›reading‹ literary texts and criticism and uncovering trends that would otherwise be prohibitively time-consuming to identify. Riddell presented several illuminating examples from his analysis on topics such as the fluctuation of scholarly interest in folk tales and the proportion of scholarship on Goethe and his works.

## Conclusion

During the lively discussions generated by these thought-provoking presentations, important questions were raised that can shape the continued development of ›distant reading‹ and its many applications to the analysis of German literature and culture in the long nineteenth century: Does evaluating data entail the same sort of close, hermeneutical reading we undertake with a literary text? What is the temporality of distant reading when computational software is involved and large amounts of data are processed? Can topic modeling identify connotations of words? Topological approaches tend to point scholars in a direction rather than provide them with answers, but are there questions that such models can answer definitively? How can we combine existing tools productively, or bring new tools into the discussion? To what extent does Google determine the corpus for us? What are the best venues for publishing work on text mining and corpus-based analysis and ensuring ongoing conversation and collaboration?

Building on and at times amending Moretti's notion of ›distant reading‹, scholars participating in Washington University's symposium offered compelling evidence that this concept and its related technologies open new and exciting possibilities for exploring German culture. Topic modeling, network analysis, and ngrams, for example, can indeed help us manage and ›read‹ the vast amount of material available online. As we have seen, publishers surveys, library records, and concepts such as ›distant reception‹ also provide new insights into nineteenth-century literature and culture. Significantly, many speakers maintained that the value of distant reading lies at least in part in how it complements or gives direction to traditional close reading. Having succeeded in demonstrating the considerable potential inherent in new digital technologies not only »to open up entirely new areas of inquiry, but also to breathe new life into some of the most venerable topics of literary studies«, the symposium represents a major contribution to literary and cultural studies and to the digital humanities.<sup>20</sup>

Lorie A. Vanchena  
University of Kansas

## Notes

<sup>1</sup> Symposium website, <http://distantreadings.wustl.edu/description/main.html> (28.08.2012).

<sup>2</sup> Cf. Luc Boltanski/Ève Chiapello, *The New Spirit of Capitalism*, trans. Gregory Elliott, London 2005; Catherine Malabou, *What Should We Do with Our Brains*, trans. Sebastian Rand, New York 2008; Wendy Hui Kyong Chun, *Programmed Visions: Software and Memory*, Cambridge, MA 2011; David Golumbia, *The Cultural Logic of Computation*, Cambridge, MA 2009.

<sup>3</sup> Cf. Alberto Manguel, *A History of Reading*, New York 2008.

<sup>4</sup> Cf. David M. Blei, Probabilistic Topics Models, *Communications of the ACM* 55:4 (2012), 77–78.

<sup>5</sup> Cf. John Burrows, ›Delta‹: A Measure of Stylistic Difference and a Guide to Likely Authorship, *Literary & Linguistic Computing* 17:3 (2002), 267–287.

<sup>6</sup> Cf. Maciej Eder/Jan Rybicki, Computational Stylistics, <https://sites.google.com/site/computationalstylistics/home> (28.08.2012).

<sup>7</sup> Culturomics website, <http://www.culturomics.org/cultural-observatory-at-harvard> (28.08.2012).

<sup>8</sup> Cf. Gao et al., Culturomics meets random fractal theory: insights into long-range correlations of social and natural phenomena over the past two centuries, *J. R. Soc. Interface*, doi: 10.1098/rsif.2011.0846.

<sup>9</sup> Cf. Voyant Tools website, <http://voyant-tools.org> (28.08.2012).

<sup>10</sup> Cf. eAqua website, <http://www.eaqua.net> (28.08.2012).

<sup>11</sup> Cf. Franco Moretti, *Graphs, Maps, Trees: Abstract Models for a Literary History*, London/New York 2005.

<sup>12</sup> Cf. Paul A. Youngman, *Black Devil and Iron Angel: The Railway in Nineteenth Century German Realism*, Washington, D.C. 2005.

<sup>13</sup> Cf. Franco Moretti, Conjectures on World Literature, *New Left Review* 1 (2000), 54–68; Stephen Best/Sharon Marcus, Surface Reading: An Introduction, *Representations* 108 (2009), 1–21; Heather Love, Close but not Deep: Literary Ethics and the Descriptive Turn, *New Literary History* 41 (2010), 371–391.

<sup>14</sup> Franco Moretti, The Slaughterhouse of Literature, *Modern Language Quarterly* 61 (2000), 207–227, cf. 207–208.

<sup>15</sup> Cf. Franco Moretti, *Graphs, Maps, Trees: Abstract Models for a Literary History*, London/New York 2005.

<sup>16</sup> Robert Darnton, Book Production in British India, 1850–1900, *Book History* 5 (2002), 239–262, quote 240.

<sup>17</sup> Cf. Robert Darnton, What is the History of Books? *Daedalus* 111:3 (1982), 65–83.

<sup>18</sup> Cf. What Middletown Read, Muncie Public Library, Center for Middletown Studies, Ball State University Library, <http://www.bsu.edu/libraries/wmr> (28.08.2012).

<sup>19</sup> Cf. Index to the Books of 1886, *Publishers' Weekly* 31 Nos. 783–4 (29 January 1887), 143.

<sup>20</sup> Symposium website, <http://distantreadings.wustl.edu/description/main.html> (28.08.2012).

2012-08-28  
JLTonline ISSN 1862-8990

**Copyright** © by the author. All rights reserved.

This work may be copied for non-profit educational use if proper credit is given to the author and JLTonline.

For other permission, please contact [JLTonline](#).

**How to cite this item:**

Lorie A. Vanchena, Reading German Culture, 1789–1918 (Conference Proceedings of: Distant Readings/Descriptive Turns: Topologies of German Culture in the Long Nineteenth Century. 21<sup>st</sup> St. Louis Symposium on German Literature & Culture, Washington University in St. Louis, March 29–31, 2012.)

In: JLTonline (28.08.2012)

Persistent Identifier: urn:nbn:de:0222-002276

Link: <http://nbn-resolving.de/urn:nbn:de:0222-002276>